Genomic Features of Transposase and Randomly Derived Recombinant CHO Clones

Steven Huhn[1], Meiping Chang[1], B. Jiang[1], Xiaoyan Tang[1], Michael Betenbaugh[2], and Zhimei Du[1]

[1]Merck & Co Inc
[2]Johns Hopkins University Department of Chemical and Biomolecular Engineering

July 18, 2022

Abstract

The use of transposase in cell line development (CLD) programs has experienced increased popularity over the past decade. However, few studies have described the mechanism of action and the genomic and phenotypic characteristics of clones derived from transposase. Additionally, how these traits impact long-term bioproduction is unknown. Here, we use chromosome painting, deep sequencing, and ddPCR to characterize the unique fingerprints associated with transposase-derived clones. Transposase reduces the cellular pool of transient vector as early as three days post transfection following transfection and expedites stable pool establishment by up to two weeks. Furthermore, , recombinant DNA expression is significantly improved up to ˜3 fold along with a greater balance of antibody heavy and light chain transcripts, resulting in higher titers in transposase generated pools. Transposase derived pools contained an often innumerable number of integration sites, representing a vast increase in integration site diversity over randomly generated pools, which were bottlenecked at 1-3 integration sites per pool. These transposase mediated integrations typically occurred in clean singlets, free of genomic scars such as deletions, inversions, and other modifications associated with legacy transfection methods which exhibited higher copy numbers per integration site. Also we observed a relative decline in gene expression with copy number increase in the randomly generated, but not the transposase derived clones. Furthermore, transposase-derived clones were more likely to exhibit enhanced a long term stability profile, including product quality attributes such as Mannose-5. This improved stability may result from circumventing mechanisms associated with the silencing of tandem repeats. Thus, transposase-mediated approaches can provide multifaceted molecular and phenotypic advantages in cell line development when compared to legacy random-integration methods.

Introduction

The transfection of exogenous DNA into a host organism is a fundamental technique in molecular biology and the first step in CHO bioproduction (Wurtele, Little, & Chartrand, 2003). A typical transfection scheme applies chemical or physical methods to permeabilize the cell membrane to nucleic acid, resulting in nuclear uptake and expression by the native transcription machinery. This process results in transient episomal expression which typically peaks around 48-72 hours post transfection and then declines thereafter with active cell divisions (Cervera, Gutierrez-Granados, Berrow, Segura, & Godia, 2015; Wurtele et al., 2003). Regulatory agencies mandate that biotherapeutic material must be derived from a single cell clone that exhibits stable expression and consistent product quality attributes (Frye et al., 2016). These features are obtained when exogenous DNA is permanently incorporated into the host genome. The stable integration of nucleic acids by traditional methods is both a rare (˜0.01% in CHO) and poorly understood event (Abraham, Tyagi, & Gottesman, 1982). Thus, the process of stable cell line generation represents a critical pain-point during cell line development (Wurtele et al., 2003).

Conventional stable CHO cell line establishment is reliant on random DNA integration via illegitimate repair

(Wurtele et al., 2003). This process is dependent on active DNA double strand breaks and subsequent recognition of foreign DNA by host DNA repair factors (Kostyrko et al., 2017; Wurtele et al., 2003). Foreign DNA is first detected by endogenous nucleases and ligases, leading to the concatemerization of many exogenous molecules (Kostyrko et al., 2017; Sargent, Brenneman, & Wilson, 1997; Wurm & Petropoulos, 1994; Wurtele et al., 2003). This process is inherently mutagenic (Wurtele et al., 2003). The ends of the concatemer are resectioned and ligated to an active DNA double strand break via the non-homologous end joining pathway (NHEJ), resulting in stable transgenic integration (Kostyrko et al., 2017; Sargent et al., 1997; Wurm & Petropoulos, 1994; Wurtele et al., 2003). In most randomized integrations these events happen once to an acceptor locus and are concurrent with chromosomal reorganization (Aeschlimann et al., 2019; Smith, 2001). The damaged acceptor locus undergoes a dynamic post-repair process that can introduce genetic changes to the integrated transgene (Sargent et al., 1997; Singer, Mesner, Friedman, Trask, & Hamlin, 2000; Sishc & Davis, 2017; Wurtele et al., 2003). Gene amplification by agents such as methotrexate accelerate such changes, and these transgenic arrays are prone to homologous recombination and epigenetic silencing (Sargent et al., 1997; Singer et al., 2000; Wurtele et al., 2003). Therefore, processes that are intrinsic to illegitimate repair by NHEJ contribute to both genetic and epigenetic instability (Chang et al., 2022; Chusainow et al., 2009; S. Huhn et al., 2022; S. C. Huhn et al., 2021).

Transposons are DNA segments that can freely change positions within the genome (Ding et al., 2005; Hickman & Dyda, 2016). Transposon systems have been widely exploited in molecular biology as an efficient tool to improve integration frequency during transfection (Ding et al., 2005). Transposon mediated integration differs from illegitimate repair in that it requires the donor sequence to be flanked with specialized motifs known as inverted terminal repeats (ITRs) (Ding et al., 2005; Mitra, Fain-Thornton, & Craig, 2008). Supplementation with an independent transposase enzyme catalyzes excision of the donor sequence from the ITR region, exposing staggered TTAA motifs with free hydroxyl groups (Mitra et al., 2008). The exposed hydroxyls then perform a nucleophilic attack on the DNA backbone of a TTAA rich acceptor locus, resulting in integration of the donor DNA onto the acceptor site (Mitra et al., 2008). Transposase mediated integration has the added benefit of preferential insertion into active chromatin (Hickman & Dyda, 2016; Tsukiyama et al., 2011). These features make transposase driven DNA integration an attractive option in cell line development pipelines. While transpose utilization has been described in terms of antibody yield and productivity, little work is available on how transposase derived clones differ from those derived from non-homologous recombination (Ahmadi et al., 2017; Matasci et al.).

Therefore, in this work we characterize the intrinsic features of CHO clones derived from transposase-based integration and compare these to those derived from randomized methods. We observe that the use of transposase dramatically shortens CLD timelines, balances expression of antibody chains, enhances recombinant DNA transcription and pool titer, and improves product quality during cell line development. Furthermore, we show that transposase lead to a marked increase in the diversity of cell integration patterns which can have benefits in terms of clone consistency. Lastly, we illustrate fundamental differences in the genomic landscape of NHEJ repaired versus transposase derived clones: while transposase mediated integrations occurred in singlets, free of genomic scars associated with legacy transfection methods, randomly generated clones were associated with changes to the host genome and transgenetic cassette. These features offer a potentially protective tool to limit the genetic silencing often observed in randomly derived clones. Taken as a whole, we demonstrate from a mechanistic perspective how the use of transposes improves cell line development platforms in a multifaceted manner.

Materials and Methods

Generation and propagation of recombinant cells

CHO hosts were grown in CD-CHO medium (Gibco, Cat: 10743029) containing 1×HT Supplement (Gibco, Cat: 11067030) and L-Glutamine (Gibco, Cat: 25030081). CHO cells were continuously cultured in shaking incubator (Kuhner) at 37°C, 5% CO^2 with or without L-glutamine.

CHO cells were transfected by electroporation with GS (glutamine synthetase) expression vector plus or

minus transposase mRNA (in titrated amounts) according to manufacturer's protocol (Neon, Thermo Fisher Scientific). The transfected cells were then selected for the expression of GS with CD-CHO media without glutamine, supplemented with 0-12.5 μM methionine sulfoximine (MSX). After continuous subculture, stable pools were cryopreserved and evaluated by fed-batch production assay. The vector system used in this manuscript utilizes a LC-GS (light chain linked to GS which also contains the IRES element) expression system, with the HC (heavy chain) being linked to a neomyicin resistance gene.

For cloning, the bulk pools were cloned into 96-well plates through limiting dilution or single cell sorting by using FACS Fusion sorter (BD). After approximately 10 to14 days, each colony was scaled to appropriately sized vessel plate.

Transposase generation

mRNAs were prepared from super transposase containing plasmid developed internally. The plasmid was first linearized, and then subjected to*In Vitro* transcription using the HiScribe T7 ARCA mRNA Kit (NEB, cat: E2060S). The mRNAs were purified using the MegaClear Kit (Thermo, Cat: AM1908).

Determination of genomic DNA copy numbers and RNA expression levels

Genomic DNAs were extracted from CHO cultures using the DNeasy Blood and Tissue Kit (QIAGEN, Germany).

For gene expression, total RNAs from CHO cells were extracted using the RNeasy Plus Mini Kit from QIAGEN (QIAGEN, Germany). cDNAs were prepared from the RNA samples by reverse transcription using the SuperScript IV VILO Master Mix (Thermo Fisher Scientific, MA).

The QX200 Droplet Digital PCR (ddPCR) System (Bio-Rad, Hercules, CA) was used to quantify the DNA copy number or the RNA transcript levels of the respective antibody chains.. Fluorescent-labeled oligo nucleotide probes for the ddPCR reactions were designed using the Primer Express Software (Applied Biosystems, Thermo Fisher Scientific, MA), and synthesized by Invitrogen (Thermo Fisher Scientific, MA).

Cytogenetics

Suspension CHO cells were treated with 0.2 μg/mL Colcemid for 4-5 hrs at 37C. Cells were then adjusted with Pre-Hypotonic Swelling Solution (Genial Helix, Cat: GGS-JL-007) and pelleted at 200g for 10 minutes. All but 0.5 mL of the remaining culture media was aspirated, and cells were resuspended in the residual media. The remaining cell volume was adjusted dropwise to 1 mL of 0.075M KCL and then to 13 mL 0.075M KCL with a serological pipette. Cultures were then incubated 20 minutes, pelleted as above, resuspended in 0.5 mL KCL, and fixed with dropwise addition of 3:1 methanol acetic acid. The cultures were then washed three times in 3:1 methanol acetic acid and aliquots of cells were dropped onto slides using a Hanabi-PVI cell-spreader. Slides were baked overnight at 50°C before any analysis. For karyotyping, only cells with non-overlapping chromosomes were considered for quantitation.

FISH

Cytogenetic spreads were first transferred to prewarmed PBS for 5-10 minutes and then transferred to 0.2 M HCL at 37C for 30 minutes. Following several washes, the spreads were digested with RNAse A (Sigma, Cat: R6148), Pepsin (Abcam, Cat: ab64201), and fixed in 1% paraformaldehyde for 10 minutes. The slides were then serially dehydrated in 80, 90 and 100% ethanol and allowed to air dry.

Slides were then denatured in water bath at 72°C for 30 minutes in 2X SSC (Lonza, Cat: 51205), allowed to cool to room temperature for 20 minutes, and immediately transferred to coplin jars containing: 0.1X SSC, 0.07 N NaOH, 0.1X SSC at 4C and 2X SSC at 4C for 1 min each. The slides were then serially dehydrated as above, warmed to 37C and denatured probe containing 1-2.5 ng/μL probe DNA was applied to the slide. The slide was covered with a coverslip, sealed with rubber cement, and incubated at 37°C for at least 18 hrs. Following hybridization, the slides were submerged in 2X SSC + 0.1% tween, the rubber cement was removed, and coverslips were floated off. Slides were washed at 72°C in 0.4X SSC for 2 minutes,

and then washed in 2X SSC + 0.1% tween for 5 minutes. Slides were then equilibrated in PBS, blocked in 0.5% Blocking Reagent (Perkin Elmer, Cat: FP1012), and incubated with anti-DIG POD (Roche, Cat: 11207733910 at 1:500) overnight. Slides were then developed using the TSA Plus Cyanine 3.5 (Perkin Elmer, Cat: NEL763001KT) exactly to manufacturer's specifications. Slides were mounted with Vectashield + DAPI (Vectashield, Cat: H-1200) before viewing on a Motorized Carl Zeiss Microscope AxioImager Z2 using ISIS software (Metasystems, Inc).

Probe Generation

1-5 µg of plasmid was digested using the Roche DIG Nick translation kit for 2-5 hours, until an approximate probe size of ˜500 bp was visualized on a 1.5% agarose gel. The probe was purified using the QIAquick Gel Extraction Kit (Qiagen, Cat: 28704 and 28706) and eluted in purified water. Prior to use, the probe was combined with 12.5 µg of salmon sperm DNA (Invitrogen, Cat: AM9680), adjusted to 0.5M NaCl and 70% ice-cold ethanol, and incubated at -20C for 2 hrs. The probe was precipitated by centrifugation at 16,000g for 1hr at 4°C, washed twice in 70% ethanol, and dissolved by incubating in 5-10 µl of 100% formamide using a thermomixer. The probe was then adjusted to 95% FISH Hybridization Buffer (Abnova, Cat: U0029), mixed, and denatured at 75°C for 10 minutes.

Chromosome Painting

Slides were prepared identically to those utilized in generic FISH, but probe was substituted for 12X Chinese Hamster Probe (Metasystems, Cat: D-1526-060-DI). Chromosome paints were visualized using ISIS software (Metasystems, Inc).

Targeted Locus Amplification

5-10 million cells were processed for Targeted Locus Amplification (TLA) according to the manufacturer's instructions (Cergentis, The Netherlands). The amount of unique integration events identified by TLA was tallied on a scaffold-by-scaffold basis per clone and is represented in heatmaps. Transpoase specific integrations were identified as sequences containing vector data extending past the ITR.

Fed Batch Production

Cells were seeded in in-house production media, glucose and lactate levels were measured daily using the RANDOX RX imola chemistry analyzer (Crumlin, UK). Cell density and viability were measured using a Beckman Coulter ViCELL cell counter (Beckman Coulter, Indianapolis, IN). mAb production levels were determined by Protein-A UPLC. In the mock perfusion assay, cells were seeded at a density of 10 million per ml. Viable cell density, mAb titer measurement, and complete media exchange was performed daily for four days.

ProA Purification

Cell-free supernatants were adjusted to 500ul in PBS and purified using PhyTips (PhyNexus, San Jose, CA; Cat: PTM 95-40-07). Briefly, tips were equilibrated in PBS, and washed in 10 mM Sodium Phosphate, 500 mM NaCl, pH 6.5 for three cycles. The captured mAb was eluted in 20 mM Sodium Acetate, pH 3.5. Protein Concentration was assessed with UV spectroscopy (Spectramax M5e, Molecular Devices, San Jose, CA) using at 260/280.

N-Glycan

Two micrograms of ProA purified material was subjected to HILIC and N-Glycan analysis using the AdvanceBio Glycan Map, Rapid Resolution (Agilent, Cat: 859700-913) column according to the manufacturer's instructions. Peak identity was resolved using Empower 3.0 software.

Statistical Analysis

Statistics (ANOVA or student's T-test) was calculated using Graphpad Prism.

Results

Transposase accelerates CLD timelines and improves pool productivity

Transposase mediated transgenic integration is a ubiquitous tool in molecular biology but is not well described in the CHO cell line development landscape. In order to better understand the differences between transposase-mediated versus random integration clones, host cell lines were transfected with circularized DNA with or without titrated amounts of transposase mRNA **(Figure 1).** At three days post transfection, cells were cultured in media to select for the expression of transgene and underwent regular passaging with complete media exchange (see Materials and Methods). We observed that the +transposase derived pools suffered little loss in viability (see red viability curve representing the +Transposase clones in **Fig. 1A**), and by passage six, the pool reached maturity with a doubling time of ˜28 hours **(red bars in Fig. 1B)** . Diluting the transposase concentration ten-fold (green) and one-hundred-fold (purple) resulted in a dose-dependent delayed recovery throughout selection (**Figure 1A,B** purple, and green curves/histograms). In these pools, doubling times did not reach maturity (˜28 hours) until passage 8 for 1:10 dilution and 10 for the 1:100 dilution. The absence of transposase (blue) lowered post-selection viabilities over all but 2 passages with stable pool establishment at 9 passages, 3 passages later than that with transposase at maximum concentration (**Figure 1B** , red bars).

Next, the pools were grown in fed batch cultures and evaluated for growth and antibody production. For a 14 day fed-batch production process, a transpose dependent increase in the production of mAb was observed **(Figure 1C,** right two graphs**).** The +transposase and 1:10 derived pools were observed to produce 11 g/L and 7 g/L of mAb, respectively **(Figure 1C,** left hand plot) while a 1:100 dilution and complete omission of transposase resulted in a striking and significant decrease in titer. To further characterize the impact of transposase implementation during stable pool development, aliquots of stable pools were collected, lysed, and extracted for total RNA and DNA every 2-3 days for up to 45 days post transfection. As shown in **Figure 1D,** the levels of cellular vector heavy chain DNA decreased dramatically over time during the transient phase of transfection (˜3 weeks), but this rate of decline was dependent on the amount of transposase. This result suggests that the transposase enzyme begins exhausting the pool of cellular plasmid immediately and perhaps prior to the first measurement, ultimately generating stable clones as early as day 3 after transfection (910 & 730 copies of heavy chain in the +transposase group [red line] vs 1680 & 1350 copies for -transposase group [blue line] on D3 and D7, respectively). A transposase dose-dependent increase of recombinant vector mRNA expression was also observed **(Figure 1E)** , especially for the antibody chains, consistent with titer results. Therefore, we surmised that transpose is most important for vector elements which are not as subject to selective pressure, given that the heavy chain was exposed to much less stringent selection in our expression platform (see Materials and Methods). Considering the expression patterns of mAb heavy and light chains at highest transposase level (**Figure 1E**), these data suggest that an added benefit of transposase incorporation is to facilitate a more balanced ratio of antibody chain expression, when compared to the random integration alternative.

Cytogenetic characterization of integration sites

We then sought to better characterize some of the cellular properties that transposase conferred to established stable pools. First, analysis by digital PCR demonstrated a progressive increase in copy number at increasing transposase concentrations during the transfection process which suggests a transposase-specific effect in driving transgene copy number (see solid bars in **Figure 2A**). We also evaluated the levels of the integrated bacterial fragment (see Bacterial Backbone [BB] in schematic of **Figure 2A** and checkered bars in the histogram), a marker for whole-vector integration through non-homologous recombination. These levels indicate the existence of random integration events in the +transposase pools, which was on average one-third of the total integrations (**Figure 2A,** compare solid and checkered histogram). Conversely, the quantity of bacterial backbone almost exactly matched the amount of heavy chain copies in the -transposase and 1:100 pools, demonstrating complete random integration of the plasmid for these cases. Next, in order to evaluate the means of integration, the pools were treated for cytogenetic analysis and probed with a cocktail of biotin labeled Bacterial backbone (a marker for random integration) and DIG labeled ITR (a marker for the transposase mediated integration) fragments **(Figure 2B&C)** . Titration of transposase

resulted in a dose dependent progressive increase in total integration sites, confirming our ddPCR results **(Figure 2 A-C).** Furthermore the +transposase pools had multiple integration foci scattered across several chromosomes **(Figure 2B)** . Both findings were in stark contrast to the 1:100 or -transposase pools, which tended to exhibit a single integration event with a high transgenic probe intensity. In these pools, the bacterial backbone always colocalized with the oft singular integration event **(Figure 2B)** . Conversely, in the +transposase or 1:10 pools, the Bacterial BB probe appeared to colocalize with only the brightest ITR foci, suggesting that these random integration sites contained a relatively higher number of copies than those resulting from transposase (**Figure 2B** , top panel, compare red arrowheads, indicating random integrants, to white arrows indicating +transposase sites). Together, these data suggest that transposase favors enhancement of transgenic insertions albeit with fewer copies per site.

The interaction of the transgene with the host chromosome plays a critical role in gene expression, and the transfection modality can play a critical role in the transgenic landscape (Wurm & Petropoulos, 1994). In order to better understand the transgene architecture with and without transposase, we utilized FISH and whole chromosome painting (mFISH) to characterize chromosomal integration sites in the transposase-derived and randomly integrated pools. We examined a 24 clones in the +transposase pool and observed that each clone demonstrated a unique configuration of integration sites among chromosomes **(Figure 3A,B)** , with up to 15 foci on one chromosome while others had only one or two (compare the zoomed images comparing Chr1 and ChrX amongst clones 1 & 2 in the lower panel of **Figure 3B**). In stark contrast, individual clones from the -transposase pools demonstrated only a narrow distribution comprised of only a few unique integrations sites throughout the population (largely consisting of just Chr5 and Chr8) for the analyzed cells **(Figure 3A,B)** . Interestingly, while integration site diversity in the -transposase pools was clearly bottlenecked, we did observed a drastic distribution of copy number at these integration sites, which could be visualized by the probe intensity at the integrated foci (compare the zoomed images comparing Chr5 amongst clones 3 & 4 in the lower panel of**Figure 3B**).

Characterization of integration sites using deep sequencing

We next sought to further elaborate on our findings using DNA sequencing. Targeted deep sequencing when married to approaches such as targeted locus amplification (TLA) allows for enhanced resolution across transgenic arrays (Aeschlimann et al., 2019; Hottentot, van Min, Splinter, & White, 2017). We were disappointed to see that conventional whole genome sequencing was not able to generate a clear and quantitative picture of pool diversity in the +transposase pools, though for the -transposase pool one major integration site could be visualized (data not shown). Therefore, to circumvent this issue, we screened and identified 19 of top producing clones from the pools (with and without transposase) identified in **Figure 3** and sequenced them using TLA (See Materials and Methods). These sequencing results closely mirrored those visualized by FISH, in which all –transposase pools demonstrated near homogeneity across a few integration sites (**Figure 4A**). In contrast, the clones derived by +transposase exhibited many different integrations sites across multiple loci**(Figure 4A).** Interestingly, the number and intensity of the heavy chain was higher than for light chain for these +transposase clones, which may be indicative of selection stringency of each chain. Additional examination of the TLA-derived genomic landscape revealed several the features associated with NHEJ in -transposase pools such as deletions (yellow arrows), inversions (blue arrows) and complex (green arrows) rearrangements (**Figure 4B,** see representative coverage plots from 3 clones, top panel). In contrast, the integration sites identified as transposase demonstrated little or no change to the surrounding chromosome and exhibited clean peaks with decreasing coverage with distance from the primer pair (**Figure 4B,** see representative coverage plots from 3 clones, bottom panel). When the copy number of these clones was analyzed by ddPCR **(Figure 5A),**the +transposase clones demonstrated a relatively higher number of copies on average. The copy number of the +transposase clones closely mirrored that of the total number of identified integration sites, suggesting that transposase tends to integrate in singlets rather than in tandem (compare histograms of the +transposase clones across**Figures 5A,B**)**.** Additionally, we detected randomly integrated (see y-axis in **Figure 5B** labeled unknown or random) DNA in +transposase clones that did not have a 1:1 relationship between copy number and the number of integration sites. This suggests that these randomly integrated sites likely contained multiple copies, in agreement for what is observed in clones

derived without transposase**.** In agreement with **Figures 2 and 3** , -transposase clones exhibited diversity in copy number only**(Figure 5A)** . Furthermore, a clear trend emerged for the subset of -transposase clones with identical integration sites; as copy number rose, gene expression tended to decline (negative correlation R=-0.75)**(Figure 5C)** . This observation is consistent with the observation that expansion of tandem arrays leads to genetic silencing (Chusainow et al., 2009; Kostyrko et al., 2017). This same trend was not evident for the +transposase clones, although these clones contained different configurations of integration sites**(Figure 5B and D)** . Taken together, these data are in agreement with the proposition that the presence of a transposase results in an increase of transgene copy number scattered across many integration sites rather than one. This spreading of integration sites may be inherently more protective against the silencing gene expression of tandem repeats, which occurs during random integration.

Assessment of cell and product quality stability in transposase and randomly derived clones

Establishment of cell lines with stable expression and consistent product quality attributes over time plays a vital role in regulatory clearance (Chusainow et al., 2009; Frye et al., 2016). In order to demonstrate stable productivity and product quality of the +transposase clones, we aged the nineteen clones in **Figure 5** to PDL75 (75 population doublings) and assessed cell productivity and percentages of Mannose-5 structures at day fourteen during fed batch culture. The specific productivity (top panel, **Figure 6**) and percentage of Mannose-5 (bottom panel, **Figure 6**) in the +transposase clones were comparable between the PDL0 and PDL75 clones for the +transposase clones (bottom panel, **Figure 6**). However, the standard deviation (Std of 8.3 for transposase clones vs 17.8 for random clones) of cell productivity was higher in the randomly derived vs the +transposase clones, demonstrating enhanced consistency in the yield of producers in the + transposase group. Additionally, amongst the -transposase clones, a greater variability in Mannose-5 was observed as the cell aged and the Mannose-5 levels were on average higher in these clones when compared to their counterparts. Thus, transposase-mediated integration does not affect stability and may provide positive benefits in terms of product quality consistency.

Discussion

In this work we contrasted the fundamental features associated with recombinant CHO clones and pools generated with and without use of transposase. We demonstrated that transposase resulted in considerably more diverse pools that are established in a shorter time frame and with an impact on cell titer and qP consistent with previous results (Kowarz, Löscher, & Marschalek, 2015) (Ahmadi et al., 2017; Matasci et al.). However, little has been done to molecularly characterize and compare clones derived from both processes. To address this question we applied molecular genetics tools for an in depth characterization of pools and clones obtained from both methodologies. Our work demonstrates that transposase derived clones contained many permutations of unique integration events which preferred the euchromatic regions of the CHO genome (S. Huhn et al., 2022). We also observed that these integrations tended to occur as clean, scattered singlets free of genomic scarring, rather than rearranged tandem repeats which can also be accompanied by deletions and inversions or other genomic events present for random integrations. Expression levels, especially of heavy chains (which exhibited weak selection), were also observed to decline with days post-transfection for those without transposase, suggesting gene silencing for random integration. Furthermore, transposase derived clones did not appear to have liability issues in terms of long-term expression, and instead offered enhancements in long-term product quality, and enhancements in clonal expression consistency.

It is well understood that transposase prefers open chromatin architecture during the integration process (Raurell-Vila, Ramos-Rodriguez, & Pasquali, 2018). Interestingly, the integration patterns for clones derived from transposase differ greatly than those observed without it. We cannot rule out that components of our vector drive the preferences observed in **Figure 3** , and furthermore, the lack of genetic diversity makes it difficult to derive commonalities in regions prone to integration by illegitimate repair in CHO. Nevertheless, it has been observed that fragile sites, AT rich regions associated with increased risk of DNA double strand breaks, are associated with increased integration by viruses and foreign DNA in human cells (Ma et al., 2012; Rassool et al., 1991). It is possible that many of loci associated with illegitimate DNA integration in CHO are associated with similar “fragile” features. Despite this, to our knowledge, no comprehensive evaluation

of these random integration patterns have been demonstrated in CHO.

Several key questions and technical challenges must be resolved in order to better utilize transposase in the context of bioproduction. Chiefly, we observed the substantial persistence of random integrations in transposase derived pools. These random sites integrated in tandem and were subject to the same liabilities as any site derived from non-homologous repair. Thus, preventing these events represent an area for improvement. In addition, we observed a poor correlation between copy number and gene expression in the transposase derived clones. These data suggest that, not all transposase sites contribute equally to gene expression. Thus future refinements are still needed in order to truly harness the capabilities of transposase for cell line development.

Acknowledgements

We would like to thank Dr. Luke Nelson for his review of this manuscript.

Declaration of interests

S. Huhn, B. Jiang, M. Chang, X. Tang, Z. Du are employees of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA and potentially own stock and/or hold stock options in Merck & Co., Inc., Kenilworth, NJ, USA. Merck & Co., Inc provided support in the form of salaries for all authors but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the "author contributions" section.

Figure Legends

Figure 1: Establishment of Recombinant CHO Pools with and without transposase

Host CHO cells were transfected with recombinant DNA in addition to transposase mRNA, with transposase mRNA diluted serially to 1:10 or 1:100, or transposase entirely omitted (-transposase). The recombinant pools were subjected to selection and passaged ten times. The viabilities are plotted in **(A)** while associated and doubling times are plotted in **(B).**The dashed black line in **(B)** represents pool maturity (about a 28-hour doubling time). Mature pools were then subjected a 14-day fed batch procedure **(C).** Titer (left panel) and qP (right panel) were measured on the indicated time points (in Days) following Protein A column purification. A red asterisk represents statistical significance ($p<0.05$) of D14 +transposase titer vs. D14 data points in other groups. **(D)** Cells were transfected with recombinant DNA with titrated transposase mRNA (blue curve without transposase, purple curve 1:100 transposase dilution, green curve 1:10 transposase dilution, and red curve undiluted transposase). Cell aliquots were then harvested, lysed, and extracted for gDNA at the indicated daily timepoints post transfection (x-axis). The DNA was then processed by ddPCR and the copy number of the antibody heavy chain was determined (y-axis). **(E)**Cell aliquots were then harvested, lysed, and extracted for mRNA at the indicated daily timepoints post transfection (bottom x-axis). The mRNA was then processed and the gene expression of the antibody recombinant chains was quantified (red curve heavy chain, blue curve light chain, green curve glutamine synthetase). The top x-axis demonstrates the category of transposase titration.

Figure 2: Copy Number and Integration Site Mediated Changes to Transposase Derived Clones and Pools

Cells were transfected with recombinant DNA with (including 1:10 or 1:100 dilutions) or without transposase mRNA. **(A)** The recombinant pools were subjected to selection, passaged, and then DNA extracted for copy number analysis with ddPCR. The average number of heavy chain copies is shown by solid bars. The average number of bacterial backbone copies (indicative of total vector incorporation) is shown by dashed bars. The number of transposase mediated copies can be estimated by the difference between the bacterial backbone and heavy chain (HC). A red asterisk represents statistical significance ($p<0.05$) of the HC copy number +transposase data point vs HC copy number in other groups. A schematic of the plasmid integration vector is shown at right. **(B)** Recombinant pools were fixed in 3:1 methanol acetic acid, spread

onto slides, and probed with FISH probes specific for the bacterial backbone (green, whole-vector integration) or the transgene (red, transposase mediated integration). White arrows indicate pure transposase mediated integrations (note the absence of co-localization with green probe), while red arrowheads mark randomized integrations where both probes colocalize. Note the intensity of random integrations (red foci intensity) compared to transposase mediated integrations, indicating higher copy number. The number of red foci per metaphase spread in +transposase, 1:10 dilution, and -transposase pools were quantified in **(C)** . Each dot represents an individual cell within the bulk pool. A red asterisk represents statistical significance ($p<0.05$) of +transposase foci number vs. all other groups, while a double red asterisk represents significance vs. 1:100 or -transposase only.

Figure 3: Integration site diversity. Recombinant CHO cells were transfected with recombinant DNA containing transposase mRNA (+Transposase) or buffer control (-Transposase). Following selection and stable pool establishment, the pools were fixed in 3:1 methanol acetic acid, dropped onto slides, and co-stained for recombinant DNA and chromosome painting. **(A)** The number of observed FISH foci in metaphase chromosome territories was quantified in heatmap form among 24 +Transposase and 36 -Transposase metaphase spreads (clones). Rows represent individual clones (in the pool) and columns represent metaphase chromosomes. As the number of observed integrations increases, heatmap coloring shifts from green (low integration number) to red (high integration number). **(B)** Representative images used in Figure 3A of a +Transposase cell (left panel) and a -Transposase clones (right panel). Chromosomes that are boxed in orange are zoomed in the bottom panel. The zoomed images demonstrate dramatically different integration patterns in the +transposase clone chromosomes 1 and X. Note the dramatic difference in fluorescence (copy number) in the -transposase clones, but not integration sites.

Figure 4: Integration Site Characteristics and Integrity in +Transposase and -Transposase derived Clones. Stable pools from transposase derived and random transfection were cloned, scaled, and subjected to fixation and transgene locus amplification (TLA).**(A)** The number of integration sites per clone was quantified and bucketed into individual genomic scaffolds; the large genomic assemblies that make up the CHO chromosomes. Columns represent individual clones and rows represent the number of integrations occurring at the indicated scaffold (genomic coordinate). As the number of observed integrations increases, heatmap coloring shifts from (low integration number) to red (high integration number). Note the completely unique composition of sites found in transposase derived clones and the scaffold/site frequently used in random integration.**(B)** The sequence read coverage profiles of loci flanking the integration sites within three +Transposase and three -Transposase clones were characterized by TLA following read mapping and bioinformatics analysis. The presence of a clean singular peak represents a locus with no obvious higher-level structural changes (see representative sites from transposase clones, boxed blue bottom panels). In the case of the clones with random integration (top red boxed panels), the sequence read coverage profiles and positions of the transgene – genome breakpoints indicate DNA structure changes including deletion (yellow arrows), inversion (blue arrows) and complex rearrangements (green arrows).

Figure 5: Relationship between Copy Number and Integration Sites in +Transposase and -Transposase derived Clones. (A) Stable pools from transposase derived and random transfection were cloned, scaled, and subjected DNA extraction and copy number analysis via ddPCR.**(A)** Histogram of DNA copy numbers (y-axis) of the heavy chain (HC) and light chain (LC) of the antibody (noted in the x-axis). Each color represents a unique clone, and the method each clone was derived is denoted under the bars on the x-axis. **(B)** Histogram of the total number of integration sites (quantified in y-axis) in each clone was bucketed into transposase mediated heavy chain specific, transposase mediated light chain specific, random, or unknown using TLA (quantified in x-axis). Each color represents a unique clone and matches**(A)** . The dashed line is drawn through both graphs at the 10 copy or 10 integration mark, respectively, and used to compare copies to integration sites. **(C)** ddPCR was used to measure the RNA expression of the heavy and light chain (barplot, left-y axis) among the random integration clones with an identical integration site and contrasted to transgene copy number (black dotted line, right y-axis). The clone ID is depicted in the x-axis. **(D)** This process was also used to measure the expression of the heavy and light antibody chains among the transposase derived clones. The correlation coefficient R between the RNA levels of the heavy chain and copy

number is indicated in the top right hand corner of the graph. (HC=Antibody heavy chain, LC=Antibody heavy chain).

Figure 6: Stability of Productivity and Antibody Characteristics in Transposase Derived and Random Integration Clones. Stable pools from transposase derived and random transfection at 0 PDL (population doubling level) or 75 PDL were grown in fed batch culture for 14 days, then cell free supernatants were collected and subjected to ProA purification and N-Glycan analysis. The top and bottom panels show the cell specific productivity, and the percentage of Mannose-5 identified in each mAb, respectively. The dotted line separates the transposase derived clones (left) and the random integration clones (right). Clones are represented in different colors (as in Figure 5) with solid colors representing PDL0 and checkered color representing PDL75. The PDL level of clones is also shown in the x-axis. Statistical significance ($p < 0.05$) between a clone's PDL0 and PDL75 trait marked by a red asterisk.

Works Cited

Abraham, I., Tyagi, J. S., & Gottesman, M. M. (1982). Transfer of genes to Chinese hamster ovary cells by DNA-mediated transformation.*Somatic Cell Genet, 8* (1), 23-39. doi:10.1007/bf01538648 [doi]

Aeschlimann, S. H., Graf, C., Mayilo, D., Lindecker, H., Urda, L., Kappes, N., . . . Laux, H. (2019). Enhanced CHO Clone Screening: Application of Targeted Locus Amplification and Next-Generation Sequencing Technologies for Cell Line Development. *Biotechnol J, 14* (7), e1800371. doi:10.1002/biot.201800371 [doi]

Ahmadi, S., Davami, F., Davoudi, N., Nematpour, F., Ahmadi, M., Ebadat, S., . . . Mahboudi, F. (2017). Monoclonal antibodies expression improvement in CHO cells by PiggyBac transposition regarding vectors ratios and design. *PLoS One, 12* (6), e0179902. doi:PONE-D-17-12959 [pii]

10.1371/journal.pone.0179902 [doi]

Cervera, L., Gutierrez-Granados, S., Berrow, N. S., Segura, M. M., & Godia, F. (2015). Extended gene expression by medium exchange and repeated transient transfection for recombinant protein production enhancement. *Biotechnol Bioeng, 112* (5), 934-946. doi:10.1002/bit.25503 [doi]

Chang, M., Kumar, A., Kumar, S., Huhn, S., Timp, W., Betenbaugh, M., & Du, Z. (2022). Epigenetic comparison of CHO hosts and clones reveals divergent methylation and transcription patterns across lineages.*Biotechnol Bioeng, 119* (4), 1062-1076. doi:10.1002/bit.28036 [doi]

Chusainow, J., Yang, Y. S., Yeo, J. H., Toh, P. C., Asvadi, P., Wong, N. S., & Yap, M. G. (2009). A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? *Biotechnol Bioeng, 102* (4), 1182-1196. doi:10.1002/bit.22158 [doi]

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., & Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell, 122* (3), 473-483. doi:S0092-8674(05)00707-5 [pii]

10.1016/j.cell.2005.07.013 [doi]

Frye, C., Deshpande, R., Estes, S., Francissen, K., Joly, J., Lubiniecki, A., . . . Anderson, K. (2016). Industry view on the relative importance of "clonality" of biopharmaceutical-producing cell lines.*Biologicals, 44* (2), 117-122. doi:S1045-1056(16)00002-6 [pii]

10.1016/j.biologicals.2016.01.001 [doi]

Hickman, A. B., & Dyda, F. (2016). DNA Transposition at Work.*Chem Rev, 116* (20), 12758-12784. doi:10.1021/acs.chemrev.6b00003 [doi]

Hottentot, Q. P., van Min, M., Splinter, E., & White, S. J. (2017). Targeted Locus Amplification and Next-Generation Sequencing.*Methods Mol Biol, 1492* , 185-196. doi:10.1007/978-1-4939-6442-0_13 [doi]

Huhn, S., Chang, M., Kumar, A., Liu, R., Jiang, B., Betenbaugh, M., . . . Du, Z. (2022). Chromosomal instability drives convergent and divergent evolution toward advantageous inherited traits in mammalian

CHO bioproduction lineages. *iScience, 25* (4), 104074. doi:S2589-0042(22)00344-3 [pii]

104074 [pii]

10.1016/j.isci.2022.104074 [doi]

Huhn, S. C., Ou, Y., Tang, X., Jiang, B., Liu, R., Lin, H., & Du, Z. (2021). Improvement of the efficiency and quality in developing a new CHO host cell line. *Biotechnol Prog* , e3185. doi:10.1002/btpr.3185 [doi]

Kostyrko, K., Neuenschwander, S., Junier, T., Regamey, A., Iseli, C., Schmid-Siegert, E., . . . Mermod, N. (2017). MAR-Mediated transgene integration into permissive chromatin and increased expression by recombination pathway engineering. *Biotechnol Bioeng, 114* (2), 384-396. doi:10.1002/bit.26086 [doi]

BIT26086 [pii]

Kowarz, E., Loscher, D., & Marschalek, R. (2015). Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines.*Biotechnol J, 10* (4), 647-653. doi:10.1002/biot.201400821 [doi]

Ma, K., Qiu, L., Mrasek, K., Zhang, J., Liehr, T., Quintana, L. G., & Li, Z. (2012). Common Fragile Sites: Genomic Hotspots of DNA Damage and Carcinogenesis. *Int J Mol Sci, 13* (9), 11974-11999. doi:10.3390/ijms130911974 [doi]

ijms-13-11974 [pii]

Matasci, M., Baldi L Fau - Hacker, D. L., Hacker Dl Fau - Wurm, F. M., Wurm, F. M., Ahmadi, S., Davami, F., . . . Mahboudi, F. A.-O. The PiggyBac transposon enhances the frequency of CHO stable cell line generation and yields recombinant lines with superior productivity and stability

Monoclonal antibodies expression improvement in CHO cells by PiggyBac transposition regarding vectors ratios and design. (1097-0290 (Electronic)).

Mitra, R., Fain-Thornton, J., & Craig, N. L. (2008). piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J, 27* (7), 1097-1109. doi:emboj200841 [pii]

10.1038/emboj.2008.41 [doi]

Rassool, F. V., McKeithan, T. W., Neilly, M. E., van Melle, E., Espinosa, R., 3rd, & Le Beau, M. M. (1991). Preferential integration of marker DNA into the chromosomal fragile site at 3p14: an approach to cloning fragile sites. *Proc Natl Acad Sci U S A, 88* (15), 6657-6661. doi:10.1073/pnas.88.15.6657 [doi]

Raurell-Vila, H., Ramos-Rodriguez, M., & Pasquali, L. (2018). Assay for Transposase Accessible Chromatin (ATAC-Seq) to Chart the Open Chromatin Landscape of Human Pancreatic Islets. *Methods Mol Biol, 1766* , 197-208. doi:10.1007/978-1-4939-7768-0_11 [doi]

Sargent, R. G., Brenneman, M. A., & Wilson, J. H. (1997). Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol Cell Biol, 17* (1), 267-277. doi:10.1128/mcb.17.1.267 [doi]

Singer, M. J., Mesner, L. D., Friedman, C. L., Trask, B. J., & Hamlin, J. L. (2000). Amplification of the human dihydrofolate reductase gene via double minutes is initiated by chromosome breaks. *Proc Natl Acad Sci U S A, 97* (14), 7921-7926. doi:10.1073/pnas.130194897 [doi]

130194897 [pii]

Sishc, B. J., & Davis, A. J. (2017). The Role of the Core Non-Homologous End Joining Factors in Carcinogenesis and Cancer.*Cancers (Basel), 9* (7). doi:10.3390/cancers9070081 [doi]

cancers-09-00081 [pii]

Smith, K. (2001). Theoretical mechanisms in targeted and random integration of transgene DNA. *Reprod Nutr Dev, 41* (6), 465-485.

Tsukiyama, T., Asano, R., Kawaguchi, T., Kim, N., Yamada, M., Minami, N., . . . Imai, H. (2011). Simple and efficient method for generation of induced pluripotent stem cells using piggyBac transposition of doxycycline-inducible factors and an EOS reporter system. *Genes Cells, 16* (7), 815-825. doi:10.1111/j.1365-2443.2011.01528.x [doi]

Wurm, F. M., & Petropoulos, C. J. (1994). Plasmid integration, amplification and cytogenetics in CHO cells: questions and comments.*Biologicals, 22* (2), 95-102. doi:S1045-1056(84)71015-3 [pii]

10.1006/biol.1994.1015 [doi]

Wurtele, H., Little, K. C., & Chartrand, P. (2003). Illegitimate DNA integration in mammalian cells. *Gene Ther, 10* (21), 1791-1799. doi:10.1038/sj.gt.3302074 [doi]

3302074 [pii]

Hosted file

`Figures_July22.pptx` available at https://authorea.com/users/495920/articles/577513-genomic-features-of-transposase-and-randomly-derived-recombinant-cho-clones